BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e135064
doi: 10.3897/biss.8.135064

OPEN ACCESS

Conference Abstract

Plinian Core: A Data Specification for Species Pages in the Real World

Francisco Pando[‡], Maria Auxiliadora Mora-Cross[§,|], Camila Plata[¶], Manuel Vargas[#], Gloria Martínez-Sagarra[¤]

‡ Real Jardin Botanico -CSIC, Madrid, Spain
§ Costa Rica Institute of Technology, Cartago, Costa Rica
| CRBio, Heredia, Costa Rica
¶ Species 2000, Bogotá, Colombia
# CRBIO, San José, Costa Rica
¤ GBIF Spain, Madrid, Spain

Corresponding author: Francisco Pando (pando@rjb.csic.es)

Received: 19 Aug 2024 | Published: 20 Aug 2024

Citation: Pando F, Mora-Cross M, Plata C, Vargas M, Martínez-Sagarra G (2024) Plinian Core: A Data Specification for Species Pages in the Real World. Biodiversity Information Science and Standards 8: e135064. https://doi.org/10.3897/biss.8.135064

Abstract

Plinian Core (PliC), a standard in development by a Biodiversity Information Standards (TDWG) task group, is a set of vocabulary terms that can be used to describe different aspects of biological species information, including all kinds of properties or traits related to taxa—biological and non-biological. PliC incorporates terms pertaining to descriptions, legal aspects, conservation, management, uses, demographics, nomenclature, or related resources (Plinian Core Task Group 2021).

Having a data specification is just a small part of what it takes to prepare and publish species information. The specification provides a structured way to describe species-related data, ensuring consistency and interoperability across different platforms and databases. Different aspects of Plinian Core and its development have been presented in the past (Pando 2018; Pando et al. 2022). In this contribution, we shift the focus out of the specification itself, to other key aspects of species information dissemination.

Plinian Core has been already put to use in various real-world scenarios. Here we select three examples to present the most significant lessons learned:

© Pando F et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

- Cross-Nature Project
  This project developed a Resource Description Framework for PliC to create an ontology-based endpoint for the Species Information System of the Spanish Ministry for the Environment (EIDOS). It provides a structured framework for describing and integrating species data following the Linked Open Data principles. The system provides information for over 85,000 species.

- Biodiversity Catalogue of SiB Colombia
  This national catalogue provides information on the natural history, conservation, threats, and biology of Colombian species. Aimed at a broad audience, it maintains scientific rigor while being accessible to anyone interested in Colombia's biodiversity. PliC has been essential in this endeavor, enabling the inclusion of both highly structured technical information and rich, engaging textual descriptions for over 6,000 species.

- EnciclоVida
  Developed by CONABIO, EnciclоVida aims to publicize the Mexican biodiversity. The platform integrates biological species information on more than 114,000 species of plants, animals, fungi, bacteria, and protozoa that CONABIO has gathered through the National Information System on Biodiversity (SNIB) in collaboration with Mexican researchers.

These implementations are based on the PliC Core abstract Model that is expressed as an XML Schema definitiion (XSD), and the supporting documentation is provided in the PliC Wiki.

Automatic aggregation of species pages is challenging and often ineffective as differences in taxonomic treatments, or in local populations (either biological or otherwise: threat status, laws and regulations, etc.) frequently create inconsistencies or even render false information. Conversely, sharing a standard across projects offers immediate benefits, such as consistency of navigation and information, or the facility for sharing software solutions.

Plinian Core terms, vocabularies and structure have been developed in collaboration with implementers and in response to real-world demands. That has resulted in a set of fairly large data specifications, but one flexible and easy to understand by domain experts who may not work comfortably with more technical approaches.

Each implementation puts emphasis on different aspects, applying different levels of granularity depending on the requirements. This is done by choosing between "unstructured" terms (for low granularity) or the "atomized classes" (for highly structured information).

For instance, dispersal information can be recorded as a text block ("Dispersal Unstructured") or using a set of specific terms: Purpose, DispersalMeans, StructureDisperse ("DispersalAtomized").

PliC reuses a number of terms from other standards (see Borrowed terms section) following a well-established best practice. Using XSD files as a basis, we link borrowed terms from external sources to their original specifications, which were also expressed as XSD files. This should provide consistency and reliability. However, this has proven to be problematic to maintain and validate, as links are broken and files changed. To keep the PliC XSD file valid, we opted for maintaining local copies of the XSD of the respective reused standards.

We also aim to present some current and future developments of PliC, especially in the context of the Living Atlases community. Thus, we will explore other elements involved in sharing species information such as the Global Biodiversity Information Facility's Integrated Publishing Toolkit extensions, SPARQL Endpoints, Frictionless Data Packages Research Object-Crates, and integration with Atlas of Living Australia modules.

Keywords

controlled vocabularies, species information, standards, use cases

Presenting author

Francisco Pando

Presented at

SPNHC-TDWG 2024

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Pando F (2018) Comparison of species information TDWG standards from the point of view of the Plinian Core specification. Biodiversity Information Science and Standards 2: 25869. https://doi.org/10.3897/biss.2.25869
- Pando F, Mora M, Ulate W, Vargas M, Plata C, Martínez-Sagarra G (2022) Plinian Core: The long and winding road. Biodiversity Information Science and Standards 6: e91584. https://doi.org/10.3897/biss.6.91584
- Plinian Core Task Group (2021) Plinian Core, a Species-level Data Specification. https://github.com/tdwg/PlinianCore. Accessed on: 2024-7-03.